Web-based Supplementary Materials for "Comparing Large Covariance Matrices under Weak Conditions on the Dependence Structure and its Application to Gene Clustering", by Jinyuan Chang, Wen Zhou, Wen-Xin Zhou and Lan Wang

This online supplementary material of the paper contains technical proofs and additional numerical results. Specifically, the notation and auxiliary lemmas are provided in Section A. Proofs for the main theoretical results in the paper are reported in Section B. Section C contains additional numerical results. Section D presents more detailed analysis of the real data example.

A Auxiliary Lemmas

A.1 Notation

First we introduce some of the basic notations. For any vector $u = (u_1, \ldots, u_p)^{\mathrm{T}} \in \mathbb{R}^p$, denote by $|u|_q$ the vector ℓ_q-norm defined by $|u|_q = \big(\sum_{k=1}^p |u_k|^q\big)^{1/q}$ for $q \geq 1$ and write $|u|_0 = \sum_{k=1}^p I(u_k \neq 0)$. For any set S, denote by S^c its complement. For a matrix $\mathbf{A} = (a_{k\ell}) \in \mathbb{R}^{p \times p}$, we denote by $\|\mathbf{A}\|_2$ the spectral norm, $\|\mathbf{A}\|_F$ the Frobenius norm, and $\|\mathbf{A}\|_1 = \sum_{k,\ell=1}^p |a_{k\ell}|$ the elementwise ℓ_1-norm. Recall that $N = n + m$ is the total sample size, $\bar{p} = p(p+1)/2$ denotes the ambient dimension and the index pairs $\{(k,\ell)\}_{1 \leq k \leq \ell \leq p}$ are organized as follows:

$$\mathcal{I}_p = \big\{(u_1, v_1), (u_2, v_2), \ldots, (u_{\bar{p}}, v_{\bar{p}})\big\} = \{(1,1), \ldots, (1,p), (2,2), \ldots, (2,p), \ldots, (p,p)\}. \tag{A.1}$$

In line with the notation in Section 2 in the paper, write $\boldsymbol{\Theta} = \boldsymbol{\Theta}_1 + nm^{-1}\boldsymbol{\Theta}_2 \in \mathbb{R}^{\bar{p} \times \bar{p}}$, where

$$\boldsymbol{\Theta}_1 = (\theta_{1,ab}), \quad \theta_{1,ab} = \mathrm{cov}\big\{(X_{u_a} - \mu_{1u_a})(X_{v_a} - \mu_{1v_a}), (X_{u_b} - \mu_{1u_b})(X_{v_b} - \mu_{1v_b})\big\} \tag{A.2}$$

and

$$\boldsymbol{\Theta}_2 = (\theta_{2,ab}), \quad \theta_{2,ab} = \mathrm{cov}\big\{(Y_{u_a} - \mu_{2u_a})(Y_{v_a} - \mu_{2v_a}), (Y_{u_b} - \mu_{2u_b})(Y_{v_b} - \mu_{2v_b})\big\}. \tag{A.3}$$

Let $\mathbf{G} = (G_1, \ldots, G_{\bar{p}})^{\mathrm{T}}$ be a $\bar{p}$-dimensional centered Gaussian random vector with covariance matrix $\mathrm{cov}(\mathbf{G}) = \mathbf{\Lambda}^{-1/2}\mathbf{\Theta}\Lambda^{-1/2}$, where $\mathbf{\Lambda} = \mathrm{diag}(\mathbf{\Theta})$.

We use C and c to denote positive constants possibly depending on K, κ, c_1 and c_2 in Assumptions (C1)–(C3) only, which may take different values at each occurrence. For two sequences of real numbers a_n and b_n, we write $a_n \asymp b_n$ if there exist two positive constants C_1, C_2 such that $C_1 \leq a_n/b_n \leq C_2$ for all $n \geq 1$, we write $a_n = O(b_n)$ if there is a positive constant C such that $a_n \leq C b_n$ holds for $n \geq 1$, and we write $a_n = o(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0$.

A.2 Auxiliary Lemmas

The following lemma is an improved Fuk-Nagaev type inequality in Banach space from Einmahl and Li (2008) (Theorem 3.1). Let $(\mathbb{B}, \|\cdot\|)$ be a real separable Banach space with topological dual $(\mathbb{B}^*, \|\cdot\|_*)$.

Lemma A.1. *Let $Z_1, \ldots, Z_n$ be independent $\mathbb{B}$-valued random variables with zero means such that $\mathbb{E}(\|Z_i\|^s) < \infty$, $1 \leq i \leq n$ for some $s \geq 2$. Then there exists a positive constant $C = C_s$ such that, for all $t > 0$,*

$$\mathbb{P}\bigg\{\bigg\|\frac{1}{n}\sum_{i=1}^n Z_i\bigg\| \geq 2\mathbb{E}\bigg(\bigg\|\frac{1}{n}\sum_{i=1}^n Z_i\bigg\|\bigg) + t\bigg\} \leq \exp\{-nt^2/(3\Lambda_n)\} + C(nt)^{-s}\sum_{i=1}^n \mathbb{E}(\|Z_i\|^s), \tag{A.4}$$

where $\Lambda_n = \sup\left\{n^{-1}\sum_{i=1}^n \mathbb{E}f^2(Z_i) : f \in \mathbb{B}^*, \|f\|_* \leq 1\right\}$.

Lemma A.1 is useful when the distributions of $\|Z_i\|$ are heavy-tailed. The next result is Theorem 6 in Delaigle et al. (2011) which provides a tight estimate on the tail probability when the underlying distributions have exponentially light tails.

Lemma A.2. *Let $Z_1, \ldots, Z_n$ be independent and identically distributed random variables with zero mean and unit variance and satisfying $\mathbb{P}(|Z_1| > t) \leq K_1 \exp(-K_2 t^\alpha)$ for all $t > 0$, where $K_1, K_2, \alpha > 0$. Write $S_n = \sum_{i=1}^n Z_i$, then there exist constants $K_3, K_4 > 0$ depending only on K_1, K_2 and α such that, for all $t > 0$,*

$$\mathbb{P}\big(|S_n| > t\big) \leq 2\exp\{-t^2/(4n)\} + K_3 \exp(-t^\alpha/K_4).$$

The following maximal inequality comes from Chernozhukov et al. (2014).

Lemma A.3. *Let $Z_1, \ldots, Z_n$ be independent $\mathbb{R}^p$-valued ($p \geq 2$) random variables. Write $Z_i = (Z_{i1}, \ldots, Z_{ip})^{\mathrm{T}}$. Then*

$$\mathbb{E}\bigg[\bigg|\sum_{i=1}^{n}\{Z_i - \mathbb{E}(Z_i)\}\bigg|_{\infty}\bigg] \leq C\big[\sigma(\log p)^{1/2} + \{\mathbb{E}(M^2)\}^{1/2}\log p\big], \tag{A.5}$$

where $C > 0$ is an absolute constant, $\sigma^2 = \max_{1\leq k\leq p}\sum_{i=1}^{n}\mathbb{E}(Z_{ik}^2)$ and $M = \max_{1\leq i\leq n}|Z_i|_{\infty}$.

The following results show that the null distribution of $\widehat{T}_{\max}$ can be approximated by that of $|\mathbf{G}|_{\infty}$ for $\mathbf{G} \sim N(\mathbf{0}, \mathbf{\Lambda}^{-1/2}\mathbf{\Theta}\mathbf{\Lambda}^{-1/2})$, which is the theoretical foundation for the validity of the proposed testing procedure.

Lemma A.4. *Suppose that Assumptions (C3) and (C4) hold.*

(i) Suppose that Assumption (C1) holds and $p \leq c_0 n^{r/2-1-\delta}$ for some $c_0, \delta > 0$. Then, under H_0,

$$\sup_{z\geq 0}\big|\mathbb{P}\big(\widehat{T}_{\max} \leq z\big) - \mathbb{P}\big(|\mathbf{G}|_{\infty} \leq z\big)\big| \leq C\big\{n^{-1/8}(\log n)^{7/8} + n^{-\delta/(r+1)}(\log n)^{3/2}\big\}. \tag{A.6}$$

(ii) Suppose that Assumption (C2) holds and that $\log p \leq c_0 n^{1/7}$ for some $c_0 > 0$. Then, under H_0,

$$\sup_{z\geq 0}\big|\mathbb{P}\big(\widehat{T}_{\max} \leq z\big) - \mathbb{P}\big(|\mathbf{G}|_{\infty} \leq z\big)\big| \leq C n^{-1/8}\{\log(pn)\}^{7/8}. \tag{A.7}$$

Here, $C > 0$ is a constant depending only on K, κ, c_0, c_1, c_2 and c.

Proof. The main idea of the proof is to first vectorize the covariance matrices and then recast the covariance testing problem as testing the equality of two mean vectors. Therefore, the arguments used in Chang et al. (2014) can be properly adapted to deal with the current setting.

By the scale and translation invariance properties of the self-normalized ratios $\hat{t}_{k\ell}$, we assume without loss of generality that $\mu_1 = \mu_2 = 0$ and $\mathrm{diag}(\mathbf{\Sigma}_1) = \mathrm{diag}(\mathbf{\Sigma}_2) = \mathbf{I}_p$. First, we introduce a new sequence of $\bar{p}$-dimensional random vectors $\{\boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{i\bar{p}})^{\mathrm{T}}\}_{i=1}^{N}$ with

$\bar{p} = p(p+1)/2$ and $N = n+m$ as follows:

$$\xi_{ia} = \begin{cases} X_{iu_a}X_{iv_a} - \sigma_{1,u_av_a}, & 1 \le i \le n, \\ -\frac{n}{m}(Y_{i-n,u_a}Y_{i-n,v_a} - \sigma_{2,u_av_a}), & n+1 \le i \le N. \end{cases}$$

It is easy to see that $\mathbb{E}(\boldsymbol{\xi}_i) = \mathbf{0}$ for $i = 1, \ldots, N$ and

$$\mathbb{C}\mathrm{ov}(\boldsymbol{\xi}_i) = \begin{cases} \Theta_1 = (\theta_{1,ab})_{1\le a,b\le\bar{p}}, & 1 \le i \le n, \\ (\frac{n}{m})^2\Theta_2 = (\frac{n}{m})^2(\theta_{2,ab})_{1\le a,b\le\bar{p}}, & n+1 \le i \le N, \end{cases}$$

where $\theta_{1,ab}$ and $\theta_{2,ab}$ are given in (A.2) and (A.3).

Define $\bar{p}\times\bar{p}$ matrices $\mathbf{D} = \mathrm{diag}(\theta_1, \ldots, \theta_{\bar{p}})$ and $\widehat{\mathbf{D}} = \mathrm{diag}(\hat{\theta}_1, \ldots, \hat{\theta}_{\bar{p}})$, where

$$\theta_a = \tfrac{n}{N}(\theta_{1,aa} + \tfrac{n}{m}\theta_{2,aa}), \quad \hat{\theta}_a = \tfrac{n}{N}(\hat{\theta}_{1,aa} + \tfrac{n}{m}\hat{\theta}_{2,aa}) = \tfrac{n}{N}(\hat{s}_{1,u_av_a} + \tfrac{n}{m}\hat{s}_{2,u_av_a}). \tag{A.8}$$

For $i = 1, \ldots, N$, write

$$\tilde{\boldsymbol{\xi}}_i = (\tilde{\xi}_{i1}, \ldots, \tilde{\xi}_{i\bar{p}})^{\mathrm{T}} = \mathbf{D}^{-1/2}\boldsymbol{\xi}_i, \quad \hat{\boldsymbol{\xi}}_i = (\hat{\xi}_{i1}, \ldots, \hat{\xi}_{i\bar{p}})^{\mathrm{T}} = \widehat{\mathbf{D}}^{-1/2}\boldsymbol{\xi}_i. \tag{A.9}$$

Under the null hypothesis $H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, the test statistic $\widehat{T}_{\max}$ can be equivalently expressed as $\widehat{T}_{\max} = \max_{1\le a\le\bar{p}} \left|N^{-1/2}\sum_{i=1}^N \hat{\xi}_{ia}\right|$. Following the proof of Theorem 1 in Chang et al. (2014), we have for any $t \ge 0$,

$$\mathbb{P}\bigg(\max_{1\le a\le\bar{p}} \bigg|N^{-1/2}\sum_{i=1}^N \hat{\xi}_{ia}\bigg| \le t\bigg) = \mathbb{P}\bigg(\max_{1\le a\le 2\bar{p}} N^{-1/2}\sum_{i=1}^N \hat{\xi}_{ia}^{\mathrm{ext}} \le t\bigg), \tag{A.10}$$

where $\{(\hat{\xi}_{i1}^{\mathrm{ext}}, \ldots, \hat{\xi}_{i\,2\bar{p}}^{\mathrm{ext}})^{\mathrm{T}}\}_{i=1}^N$ is a sequence of $2\bar{p}$-dimensional random vectors defined by $\hat{\xi}_{ia}^{\mathrm{ext}} = \hat{\xi}_{ia}$ for $1 \le a \le \bar{p}$ and $\hat{\xi}_{ia}^{\mathrm{ext}} = -\hat{\xi}_{ia}$ for $\bar{p}+1 \le a \le 2\bar{p}$. Without loss of generality, it suffices to prove the result for

$$\widehat{T}^{+} = \max_{1\le a\le\bar{p}} N^{-1/2}\sum_{i=1}^N \hat{\xi}_{ia}. \tag{A.11}$$

Case 1 (Polynomial decay). Suppose that Assumption (C1) holds and $p \le c_0 n^{r/2-1-\delta}$ for some $c_0, \delta > 0$ so that $\alpha_{n,p} = pn^{1-r/2} \le c_0 n^{-\delta}$. Define an intermediate approximation

$$\widetilde{T}^+ = \max_{1\le a\le \bar{p}} N^{-1/2} \sum_{i=1}^{N} \tilde{\xi}_{ia}$$

for $\tilde{\xi}_{ia}$ as in (A.9) and note that

$$\big|\widehat{T}^+ - \widetilde{T}^+\big| \le \max_{1\le a\le \bar{p}} \big|(\hat{\theta}_a/\theta_a)^{1/2} - 1\big| \cdot \widehat{T}_{\max}. \tag{A.12}$$

For every $(k,\ell) \in \mathcal{I}_p$, put $\tilde{s}_{1,k\ell} = n^{-1}\sum_{i=1}^n (X_{ik}X_{i\ell} - \sigma_{1,k\ell})^2$ and $\tilde{s}_{2,k\ell} = m^{-1}\sum_{j=1}^m (Y_{jk}Y_{j\ell} - \sigma_{2,k\ell})^2$ so that $\mathbb{E}(\tilde{s}_{1,k\ell}) = s_{1,k\ell}$ and $\mathbb{E}(\tilde{s}_{2,k\ell}) = s_{2,k\ell}$. To upper bound $\widehat{T}_{\max}$, first we show that $\tilde{s}_{\nu,k\ell}$ is the dominating term of $\hat{s}_{\nu,k\ell}$ for $\nu = 1,2$. Using arguments similar to those in the proof of Lemma A.5 yields that, under Assumption (C1),

$$\mathbb{P}\bigg\{ \max_{(k,\ell)\in\mathcal{I}_p} |\hat{s}_{\nu,k\ell} - \tilde{s}_{\nu,k\ell}| \ge C\alpha_{n,p}^{2/(r+2)} \bigg\} \le C\alpha_{n,p}^{2/(r+2)}. \tag{A.13}$$

Moreover, it follows from Theorem 2.19 in de la Peña et al. (2009) that for any $\varepsilon > 0$,

$$\mathbb{P}\big(\tilde{s}_{1,k\ell} \le s_{1,k\ell} - \varepsilon\big) \le \exp(-n\varepsilon^2/[2\mathbb{E}\{(X_kX_\ell - \sigma_{1,k\ell})^4\}]).$$

This implies by taking $\varepsilon = \varepsilon_{n,p} \asymp n^{-1/2}\{\log(pn)\}^{1/2}\max_{1\le k\le p}(\mathbb{E}X_k^8)^{1/2}$ and the union bound that

$$\mathbb{P}\big[\tilde{s}_{1,k\ell} \le s_{1,k\ell} - Cn^{-1/2}\{\log(pn)\}^{1/2} \text{ for some } (k,\ell)\in\mathcal{I}_p\big] \le Cn^{-1}. \tag{A.14}$$

Under Assumptions (C1), (C3) and (C4), it follows from (A.13), (A.14) and Theorem 2.16 in de la Peña et al. (2009) that, for any $t > 0$ and for all sufficiently large n,

$$\begin{aligned}
&\mathbb{P}\big(\widehat{T}_{\max} \ge t\big) \\
&\le \sum_{a=1}^{\bar{p}} \mathbb{P}\bigg\{\bigg|\sum_{i=1}^{N}\xi_{ia}\bigg| \ge ct\sqrt{n}(\tilde{s}_{1,u_av_a} + s_{1,u_av_a} + \tilde{s}_{2,u_av_a} + s_{2,u_av_a})^{1/2}\bigg\} + C\big\{\alpha_{n,p}^{2/(r+2)} + n^{-1}\big\} \\
&\le \sum_{a=1}^{\bar{p}} \mathbb{P}\Bigg[\bigg|\sum_{i=1}^{N}\xi_{ia}\bigg| \ge ct\bigg\{\sum_{i=1}^{N}\xi_{ia}^2 + \sum_{i=1}^{N}\mathbb{E}(\xi_{ia}^2)\bigg\}^{1/2}\Bigg] + C\big\{\alpha_{n,p}^{2/(r+2)} + n^{-1}\big\} \\
&\le C\big\{\bar{p}\exp(-ct^2) + \alpha_{n,p}^{2/(r+2)}\big\}.
\end{aligned} \tag{A.15}$$

Recall that $\mathbf{G} = (G_1, \ldots, G_{\bar{p}})^{\mathrm{T}}$ is a $\bar{p}$-dimensional centered Gaussian random vector with covariance matrix $\mathbb{E}(\mathbf{G}\mathbf{G}^{\mathrm{T}}) = N^{-1}\sum_{i=1}^{N}\mathbb{E}(\tilde{\xi}_i\tilde{\xi}_i^{\mathrm{T}}) = \mathbf{\Lambda}^{-1/2}\mathbf{\Theta}\mathbf{\Lambda}^{-1/2}$, where $\mathbf{\Lambda} = \mathrm{diag}(\mathbf{\Theta})$. Let

$$W_0^+ = \max_{1\le k\le \bar{p}} G_k. \tag{A.16}$$

Then applying Theorem 2.2 and Lemma 2.2 in Chernozhukov et al. (2013) gives, for every $t \in (0,1)$,

$$\begin{aligned}
&\sup_{z\in\mathbb{R}}\big|\mathbb{P}\big(\widetilde{T}^+ \le z\big) - \mathbb{P}\big(W_0^+ \le z\big)\big| \\
&\lesssim n^{-1/8}\{\log(pn/t)\}^{7/8}\max_{1\le k\le p}\{\mathbb{E}(X_k^8)\}^{1/4} + t^{-1/r}\alpha_{n,p}^{1/r}\{\log(pn/t)\}^{3/2}\max_{1\le k\le p}\{\mathbb{E}(X_k^{2r})\}^{1/r} + t.
\end{aligned}$$

This implies by taking $t = \alpha_{n,p}^{1/(r+1)}$ that

$$\sup_{z\in\mathbb{R}}\big|\mathbb{P}\big(\widetilde{T}^+ \le z\big) - \mathbb{P}\big(W_0^+ \le z\big)\big| \lesssim n^{-1/8}(\log n)^{7/8} + \alpha_{n,p}^{1/(r+1)}(\log n)^{3/2}. \tag{A.17}$$

Note that $G_1, \ldots, G_{\bar{p}}$ are standard normal random variables. By (A.16) and Theorem 3 in Chernozhukov et al. (2014), we have for every $\varepsilon > 0$,

$$\sup_{z\in\mathbb{R}}\mathbb{P}\big(|W_0^+ - z| \le \varepsilon\big) \le 4\varepsilon\big\{1 + 2(\log p)^{1/2}\big\}. \tag{A.18}$$

Together, (A.12), (A.15), (A.17), (A.18) and Lemma A.5 yield, for every $z \in \mathbb{R}$ and $\varepsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\big(\widehat{T}^+ \le z\big) \\
&\le \mathbb{P}\big(\widetilde{T}^+ \le z + \varepsilon\big) + \mathbb{P}\big(|\widehat{T}^+ - \widetilde{T}^+| > \varepsilon\big) \\
&\le \mathbb{P}\big(W_0^+ \le z + \varepsilon\big) + \mathbb{P}\big\{\widehat{T}_{\max} > c\,\varepsilon\alpha_{n,p}^{-2/(r+2)}\big\} + C\big\{n^{-1/8}(\log n)^{7/8} + \alpha_{n,p}^{1/(r+1)}(\log n)^{3/2}\big\} \\
&\le \mathbb{P}\big(W_0^+ \le z\big) + C\big[\varepsilon(\log p)^{1/2} + p^2\exp\{-c\,\varepsilon^2\alpha_{n,p}^{-4/(r+2)}\} + n^{-1/8}(\log n)^{7/8} + \alpha_{n,p}^{1/(r+1)}(\log n)^{3/2}\big].
\end{aligned}$$

Taking $\varepsilon = \varepsilon_{n,p} \asymp \alpha_{n,p}^{2/(r+2)}(\log n)^{1/2}$, this implies that

$$\mathbb{P}\big(\widehat{T}^+ \le z\big) \le \mathbb{P}\big(W_0^+ \le z\big) + C\big\{n^{-1/8}(\log n)^{7/8} + \alpha_{n,p}^{1/(r+1)}(\log n)^{3/2}\big\}. \tag{A.19}$$

A similar argument leads to the reverse inequality and completes the proof of (A.6).

Case 2 (Exponential decay). Suppose that Assumption (C2) holds and that $\log p \le c_0 n^{1/7}$. Similar to (A.13) and (A.17), now we have

$$\mathbb{P}\bigg(\max_{(k,\ell)\in\mathcal{I}_p} |\hat{s}_{\nu,k\ell} - \tilde{s}_{\nu,k\ell}| \ge C \max\big[n^{-1/2}\{\log(pn)\}^{1/2}, n^{-1}\{\log(pn)\}^2\big]\bigg) \le Cn^{-1}$$

and

$$\sup_{z\in\mathbb{R}} \big|\mathbb{P}\big(\widehat{T}^+ \le z\big) - \mathbb{P}\big(W_0^+ \le z\big)\big| \le C\big[n^{-1/8}\{\log(pn)\}^{7/8} + n^{-1/2}\{\log(pn)\}^{5/2}\big].$$

Arguments similar to those employed for the Case 1 can be used the prove (A.7). $\square$

For any fixed $\alpha \in (0,1)$, our main purpose is to study the asymptotic properties of the data-driven test $\Psi_\alpha = I\{\widehat{T}_{\max} > \hat{c}_\alpha\}$ for $\hat{c}_\alpha$ given in Section 2.2. At this point, Lemma A.4 is not directly useful as the distribution of $|\mathbf{G}|_\infty$ is unknown. By definition, the multiplier bootstrap statistic $\widehat{T}^\dagger_{\max}$ given in (2.4) can be written as

$$\widehat{T}^\dagger_{\max} = |\mathbf{G}^*|_\infty = \max_{1\le a\le \bar{p}} |G^*_a|,$$

where $\mathbf{G}^* = (G^*_1,\ldots,G^*_{\bar{p}})^{\mathrm{T}} \in \mathbb{R}^{\bar{p}}$ with $G^*_a = \hat{t}^\dagger_{u_a v_a}$ for $(u_a, v_a) \in \mathcal{I}_p$ as in (A.1). Conditional on $\mathcal{X}_n$ and $\mathcal{Y}_m$, $\mathbf{G}^*$ has the centered Gaussian distribution with covariance matrix $\widehat{\mathbf{\Lambda}}^{-1/2}\widehat{\mathbf{\Theta}}\widehat{\mathbf{\Lambda}}^{-1/2}$, where $\widehat{\mathbf{\Lambda}} = \mathrm{diag}(\widehat{\mathbf{\Theta}})$, $\widehat{\mathbf{\Theta}} = \widehat{\mathbf{\Theta}}_1 + nm^{-1}\widehat{\mathbf{\Theta}}_2$, $\widehat{\mathbf{\Theta}}_1 = (\hat{\theta}_{1,ab})$ and $\widehat{\mathbf{\Theta}}_2 = (\hat{\theta}_{2,ab})$ with

$$\hat{\theta}_{1,ab} = \frac{1}{n}\sum_{i=1}^n \big\{(X_{iu_a} - \bar{X}_{u_a})(X_{iv_a} - \bar{X}_{v_a}) - \hat{\sigma}_{1,u_a v_a}\big\}\big\{(X_{iu_b} - \bar{X}_{u_b})(X_{iv_b} - \bar{X}_{v_b}) - \hat{\sigma}_{1,u_b v_b}\big\},$$
$$\hat{\theta}_{2,ab} = \frac{1}{m}\sum_{j=1}^m \big\{(Y_{ju_a} - \bar{Y}_{u_a})(Y_{jv_a} - \bar{Y}_{v_a}) - \hat{\sigma}_{2,u_a v_a}\big\}\big\{(Y_{ju_b} - \bar{Y}_{u_b})(Y_{jv_b} - \bar{Y}_{v_b}) - \hat{\sigma}_{2,u_b v_b}\big\}.$$

Note that $\hat{\theta}_{1,ab}$ and $\hat{\theta}_{2,ab}$ are consistent estimates, respectively, of $\theta_{1,ab}$ and $\theta_{2,ab}$ given in (A.2) and (A.3). In particular, $\hat{\theta}_{1,aa} = \hat{s}_{1,u_a v_a}$ and $\hat{\theta}_{2,aa} = \hat{s}_{2,u_a v_a}$. For $(k,\ell) \in \mathcal{I}_p$, recall that $s_{1,k\ell}$ and $s_{2,k\ell}$ are the the variances of $(X_k - \mu_{1k})(X_\ell - \mu_{1\ell})$ and $(Y_k - \mu_{2k})(Y_\ell - \mu_{2\ell})$, respectively. Lemmas A.5 and A.6 below are on the concentration inequalities for $\max_{(k,\ell)\in\mathcal{I}_p} |\hat{\theta}_{\nu,k\ell} - \theta_{\nu,k\ell}|$ and $\max_{(k,\ell)\in\mathcal{I}_p} |\hat{s}_{\nu,k\ell} - s_{\nu,k\ell}|$ with $\nu = 1,2$.

Lemma A.5. *Assume that the conditions of Lemma A.4 hold.*

(i) Under Assumption (C1), there exists a constant $C > 0$ depending only on K such that

$$\mathbb{P}\bigg(\max_{(k,\ell)\in\mathcal{I}_p} \frac{|\hat{s}_{1,k\ell} - s_{1,k\ell}|}{\sigma_{1,kk}\sigma_{1,\ell\ell}} \geq C\big[n^{-1/2}\{\log(pn)\}^{1/2} + \alpha_{n,p}^{2/(r+2)} + \alpha_{n,p}^{2/r}\log p\big] \bigg) \leq C\alpha_{n,p}^{2/(r+2)}, \tag{A.20}$$

where $\alpha_{n,p} = pn^{1-r/2}$ for $r \geq 4$ as in Assumption (C1).

(ii) Under Assumption (C2), there exists some constant $C > 0$ depending only on K and κ such that

$$\mathbb{P}\bigg(\max_{(k,\ell)\in\mathcal{I}_p} \frac{|\hat{s}_{1,k\ell} - s_{1,k\ell}|}{\sigma_{1,kk}\sigma_{1,\ell\ell}} \geq C\max\big[n^{-1/2}\{\log(pn)\}^{1/2}, n^{-1}\{\log(pn)\}^{2}\big] \bigg) \leq Cn^{-1}. \tag{A.21}$$

Similar results hold for $\max_{(k,\ell)\in\mathcal{I}_p} |\hat{s}_{2,k\ell} - s_{2,k\ell}|/(\sigma_{2,kk}\sigma_{2,\ell\ell})$ with n replaced by m.

Proof. By the translation invariance property of $\hat{s}_{1,k\ell}$, we assume without loss of generality that $\mathbb{E}(X) = \mathbb{E}(Y) = 0$ and $\mathrm{diag}(\mathbf{\Sigma}_1) = \mathrm{diag}(\mathbf{\Sigma}_2) = \mathbf{I}_p$. Direct calculations show that

$$\begin{aligned}\hat{s}_{1,k\ell} &= \frac{1}{n}\sum_{i=1}^{n}(X_{ik}X_{i\ell})^2 - \bar{V}_{k\ell}^2 + 6\bar{X}_k\bar{X}_\ell V_{k\ell} \\ &\quad - 2\bar{X}_k\frac{1}{n}\sum_{i=1}^{n}X_{ik}X_{i\ell}^2 - 2\bar{X}_\ell\frac{1}{n}\sum_{i=1}^{n}X_{ik}^2X_{i\ell} + \bar{X}_k^2\bar{V}_{\ell\ell} + \bar{V}_{kk}\bar{X}_\ell^2 - 4\bar{X}_k^2\bar{X}_\ell^2,\end{aligned} \tag{A.22}$$

where $\bar{V}_{k\ell} = n^{-1}\sum_{i=1}^{n} X_{ik}X_{i\ell}$. Define

$$\begin{aligned}\Delta_0 &= \max_{1\leq k\leq p} |\bar{X}_k|, \\ \Delta_1 &= \max_{(k,\ell)\in\mathcal{I}_p} \big|V_{k\ell} - \sigma_{1,k\ell}\big|, \\ \Delta_2 &= \max_{1\leq k,\ell\leq p} \bigg|\frac{1}{n}\sum_{i=1}^{n} X_{ik}X_{i\ell}^2 - \mathbb{E}(X_{ik}X_{i\ell}^2)\bigg|, \text{ and} \\ \Delta_3 &= \max_{(k,\ell)\in\mathcal{I}_p} \bigg|\frac{1}{n}\sum_{i=1}^{n}(X_{ik}X_{i\ell})^2 - \mathbb{E}\{(X_{ik}X_{i\ell})^2\}\bigg|.\end{aligned}$$

Using these notations, our goal is to bound the random quantities $\{\Delta_\nu\}_{\nu=0,1,2,3}$. The main strategy is to first show concentration of Δ_ν around its expectation by Lemma A.1 and then upper bound the expectation via the maximal inequality (A.5). We restrict attention to Δ_1 and Δ_3 because the proof for Δ_0 and Δ_2 is almost identical.

Case 1 (Polynomial decay). Suppose that Assumption (C1) holds. First we study Δ_1. In Lemma A.1, let $\mathbb{B} = \mathbb{R}^{\bar p}$ with $\bar p = p(p+1)/2$, $s = r$ and $\|\cdot\| = |\cdot|_\infty$ so that $\|\cdot\|_* = |\cdot|_1$. Note that, for any $f = (f_1, \ldots, f_{\bar p})^{\mathrm{T}} (\neq 0) \in \mathbb{R}^{\bar p}$ with $|f|_1 = \sum_{k=1}^{\bar p} |f_k| \le 1$, it follows from Jensen's inequality that

$$f^2(Z_i) = \bigg(\sum_{k=1}^{\bar p} f_k Z_{ik} \bigg)^2 \le \bigg(\frac{\sum_{k=1}^{\bar p} |f_k||Z_{ik}|}{\sum_{k=1}^{\bar p} |f_k|} \bigg)^2 \le |f|_1^{-1} \sum_{k=1}^{\bar p} |f_k| Z_{ik}^2.$$

This implies by definition that $\Lambda_n \le \max_{(k,\ell)\in\mathcal{I}_p} n^{-1} \sum_{i=1}^n \mathbb{E}\{(X_{ik}X_{i\ell})^2\} \le \max_{1\le k\le p} \mathbb{E}(X_k^4)$. Then we have for any $t > 0$,

$$\mathbb{P}\big\{\Delta_1 \ge 2\mathbb{E}(\Delta_1) + t\big\} \le \exp[-(nt^2)/\{3 \max_k \mathbb{E}(X_k^4)\}] + C\, t^{-r} n^{1-r}\, \mathbb{E}(|X|_\infty^{2r}). \tag{A.23}$$

By Lemma A.3,

$$\mathbb{E}(\Delta_1) \lesssim n^{-1/2} (\log p)^{1/2} \max_{1\le k\le p} \{\mathbb{E}(X_k^4)\}^{1/2} + n^{-1} \log p \left\{ \mathbb{E}\bigg(\max_{1\le i\le n} \max_{(k,\ell)\in\mathcal{I}_p} X_{ik}^2 X_{i\ell}^2 \bigg) \right\}^{1/2}. \tag{A.24}$$

In view of (A.23), (A.24) and the inequalities $\mathbb{E}(|X|_\infty^{2r}) \le p \max_{1\le k\le p} \mathbb{E}(X_k^{2r})$ and

$$\mathbb{E}\bigg(\max_{1\le i\le n} \max_{(k,\ell)\in\mathcal{I}_p} X_{ik}^2 X_{i\ell}^2 \bigg) \le \left\{ \mathbb{E}\bigg(\max_{1\le i\le n} \max_{1\le k\le p} X_{ik}^{2r} \bigg) \right\}^{2/r} \le (pn)^{2/r} \max_{1\le k\le p} \{\mathbb{E}(X_k^{2r})\}^{2/r}$$

we take

$$t = t_{n,p} \asymp \max\left[n^{-1/2} \{\log(pn)\}^{1/2} \max_{1\le k\le p} \{\mathbb{E}(X_k^4)\}^{1/2},\ \beta_{n,p}^{1/(r+1)} \max_{1\le k\le p} \{\mathbb{E}(X_k^{2r})\}^{1/r} \right]$$

with $\beta_{n,p} = pn^{1-r}$ such that, with probability at least $1 - C\beta_{n,p}^{1/(r+1)}$,

$$\Delta_1 \lesssim n^{-1/2} \{\log(pn)\}^{1/2} + \beta_{n,p}^{1/r} \log p + \beta_{n,p}^{1/(r+1)}. \tag{A.25}$$

For Δ_3, using Lemma A.1 with $s = r/2$ and Lemma A.3, successively yields

$$\mathbb{P}\big\{\Delta_3 \ge 2\mathbb{E}(\Delta_3) + t\big\} \le \exp[-(nt^2)/\{3 \max_k \mathbb{E}(X_k^8)\}] + Ct^{-r/2} pn^{1-r/2} \max_{1\le k\le p} \mathbb{E}(X_k^{2r})$$

for all $t > 0$ and

$$\mathbb{E}(\Delta_3) \lesssim n^{-1/2} \max_{1\le k\le p} \{\mathbb{E}(X_k^8)\}^{1/2} + p^{2/r} n^{2/r-1} \log p \max_{1\le k\le p} \{\mathbb{E}(X_k^{2r})\}^{2/r}.$$

In particular, letting

$$t = t_{n,p} \asymp \max\left[n^{-1/2}\{\log(pn)\}^{1/2} \max_k \{\mathbb{E}(X_k^8)\}^{1/2},\ \alpha_{n,p}^{2/(r+2)} \max_k \{\mathbb{E}(X_k^{2r})\}^{2/r}\right]$$

with $\alpha_{n,p} = pn^{1-r/2}$ gives

$$\mathbb{P}\Big(\Delta_3 \ge C\big[n^{-1/2}\{\log(pn)\}^{1/2} + \alpha_{n,p}^{2/(r+2)} + \alpha_{n,p}^{2/r} \log p\big]\Big) \lesssim \alpha_{n,p}^{2/(r+2)}. \tag{A.26}$$

Together, (A.22), (A.25) and (A.26) proves (A.20).

Case 2 (Exponential decay). Suppose that Assumption (C2) holds. Then for every $(k, \ell) \in \mathcal{I}_p$ and $t > 0$,

$$\mathbb{E}(e^{t|X_k X_\ell|}) \le \mathbb{E}\{e^{t(X_k^2+X_\ell^2)/2}\} \le \{\mathbb{E}(e^{tX_k^2})\}^{1/2}\{\mathbb{E}(e^{tX_\ell^2})\}^{1/2}.$$

Applying Lemma A.2 with $\alpha = 1/2$ yields that, for any $t > 0$,

$$\begin{aligned}
&\mathbb{P}\Big(\Delta_3 \ge \max_{1\le k\le \ell\le p} [\mathbb{E}\{(X_k X_\ell)^4\}]^{1/2} \cdot t\Big) \\
&\le \sum_{(k,\ell)\in\mathcal{I}_p} \mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^n (X_{ik}X_{i\ell})^2 - E\{(X_k X_\ell)^2\}\right| \ge \mathbb{V}\mathrm{ar}^{1/2}\{(X_k X_\ell)^2\} \cdot t\right] \\
&\le \sum_{(k,\ell)\in\mathcal{I}_p} \big[2\exp(-nt^2/4) + C\exp\{-c(nt)^{1/2}\}\big].
\end{aligned}$$

This implies by taking $t = t_{n,p} \asymp \max\big[n^{-1/2}\{\log(pn)\}^{1/2}, n^{-1}\{\log(pn)\}^2\big]$ that

$$\mathbb{P}\Big(\Delta_3 \ge C\max\big[n^{-1/2}\{\log(pn)\}^{1/2}, n^{-1}\{\log(pn)\}^2\big]\Big) \lesssim n^{-1}. \tag{A.27}$$

Similar arguments can be employed to prove (A.27) for Δ_0, Δ_1 and Δ_2, which completes the proof of (A.21). $\square$

Lemma A.6. *Assume that the conditions of Lemma A.5 hold.*

(i) Under Assumption (C1), there exists a constant $C > 0$ depending only on K such that

$$\mathbb{P}\bigg(\max_{1\le a\le b\le \bar{p}} \frac{|\hat{\theta}_{1,ab}-\theta_{1,ab}|}{\sigma_{1,u_av_a}\sigma_{1,u_bv_b}} \ge C\big[n^{-1/2}\{\log(pn)\}^{1/2} + \alpha_{n,p}^{2/(r+2)} + \alpha_{n,p}^{2/r}\log p\big]\bigg) \le C\alpha_{n,p}^{2/(r+2)}.$$

(ii) Under Assumption (C2), there exists some constant $C > 0$ depending only on K and κ such that

$$\mathbb{P}\bigg(\max_{1\le a\le b\le \bar{p}} \frac{|\hat{\theta}_{1,ab}-\theta_{1,ab}|}{\sigma_{1,u_av_a}\sigma_{1,u_bv_b}} \ge C\max\big[n^{-1/2}\{\log(pn)\}^{1/2}, n^{-1}\{\log(pn)\}^2\big]\bigg) \le Cn^{-1}.$$

Similar results hold for $\max_{1\le a\le b\le \bar{p}}|\hat{\theta}_{2,ab}-\theta_{2,ab}|/(\sigma_{2,u_av_a}\sigma_{2,u_bv_b})$ with n replaced by m.

The proof of Lemma A.6 is essentially same to that of Lemma A.5 and is omitted.

B Proof of the Main Theorems

B.1 Proof of Theorem 1

Let $\mathbf{\Sigma}_0 = (\sigma_{0,ab})_{1\le a,b\le \bar{p}}$ be the covariance matrix of $\mathbf{G}$ and let $\widehat{\mathbf{\Sigma}}_0 = (\hat{\sigma}_{0,ab})_{1\le a,b\le \bar{p}}$ be the covariance matrix of $\mathbf{G}^*$ conditional on $\mathcal{X}_n$ and $\mathcal{Y}_m$. Note that $\sigma_{0,aa} = \hat{\sigma}_{0,aa} = 1$ for $a = 1,\ldots,\bar{p}$. Define

$$W_0 = \max_{1\le k\le \bar{p}}|G_k| = \max_{1\le k\le \bar{p}}\max(G_k,-G_k), \quad W_0^* = \max_{1\le k\le \bar{p}}|G_k^*| = \max_{1\le k\le \bar{p}}\max(G_k^*,-G_k^*). \tag{B.1}$$

Applying Theorem 2 in Chernozhukov et al. (2014) to the $(2\bar{p})$-dimensional random vectors $\bar{\mathbf{G}} = (G_1,\ldots,G_{\bar{p}},-G_1,\ldots,-G_{\bar{p}})^{\mathrm{T}}$ and $\bar{\mathbf{G}}^* = (G_1^*,\ldots,G_{\bar{p}}^*,-G_1^*,\ldots,-G_{\bar{p}}^*)^{\mathrm{T}}$ gives

$$\sup_{z\ge 0}\big|\mathbb{P}\big(|\mathbf{G}|_\infty \le z\big) - \mathbb{P}_g\big(|\mathbf{G}^*|_\infty\le z\big)\big| \le C\Delta_0^{1/3}(\log p)^{1/3}\{(\log p)\vee\log(1/\Delta_0)\}^{1/3},$$

where $C > 0$ is an absolute constant and $\Delta_0 = \max_{1\le a<b\le\bar{p}}|\sigma_{0,ab}-\hat{\sigma}_{0,ab}|$. Under the conditions of Theorem 1, this implies by Lemma A.6 that as $n,m\to\infty$,

$$\sup_{z\ge 0}\big|\mathbb{P}\big(|\mathbf{G}|_\infty \le z\big) - \mathbb{P}_g\big(|\mathbf{G}^*|_\infty\le z\big)\big| \xrightarrow{P} 0. \tag{B.2}$$

Together, the definition of c_α, (B.2) and Lemma A.4 imply

$$
\begin{aligned}
&\mathbb{P}_{H_0}\big(\Psi_\alpha = 1\big) - \alpha \\
&= 1 - \alpha - \mathbb{P}_{H_0}\big(\widehat{T}_{\max} \le \hat{c}_\alpha\big) \\
&\le 1 - \alpha - \mathbb{P}\big(|\mathbf{G}|_\infty \le \hat{c}_\alpha\big) + \sup_{z\ge 0}\big|\mathbb{P}\big(|\mathbf{G}|_\infty \le z\big) - \mathbb{P}_{H_0}\big(\widehat{T}_{\max} \le z\big)\big| \\
&\le 1 - \alpha - \mathbb{P}_g\big(|\mathbf{G}^*|_\infty \le \hat{c}_\alpha\big) \\
&\qquad + \sup_{z\ge 0}\big|\mathbb{P}_g\big(|\mathbf{G}^*|_\infty \le z\big) - \mathbb{P}\big(|\mathbf{G}|_\infty \le z\big)\big| + \sup_{z\ge 0}\big|\mathbb{P}\big(|\mathbf{G}|_\infty \le z\big) - \mathbb{P}_{H_0}\big(\widehat{T}_{\max} \le z\big)\big| \\
&\le \sup_{z\ge 0}\big|\mathbb{P}_g\big(|\mathbf{G}^*|_\infty \le z\big) - \mathbb{P}\big(|\mathbf{G}|_\infty \le z\big)\big| + \sup_{z\ge 0}\big|\mathbb{P}\big(|\mathbf{G}|_\infty \le z\big) - \mathbb{P}_{H_0}\big(\widehat{T}_{\max} \le z\big)\big|.
\end{aligned}
$$

For the lower bound, it follows from (A.18) that

$$
\begin{aligned}
&\mathbb{P}_{H_0}\big(\Psi_\alpha = 1\big) - \alpha \\
&\ge \mathbb{P}\big(|\mathbf{G}|_\infty > \hat{c}_\alpha\big) - \alpha - \sup_{z\ge 0}\big|\mathbb{P}\big(|\mathbf{G}|_\infty \le z\big) - \mathbb{P}_{H_0}\big(\widehat{T}_{\max} \le z\big)\big| \\
&\ge \mathbb{P}\big(|\mathbf{G}|_\infty > \hat{c}_\alpha - n^{-1}\big) - \alpha - \sup_{z\ge 0}\big|\mathbb{P}\big(|\mathbf{G}|_\infty \le z\big) - \mathbb{P}_{H_0}\big(\widehat{T}_{\max} \le z\big)\big| - Cn^{-1}(\log p)^{1/2} \\
&\ge \mathbb{P}_g\big(|\mathbf{G}^*|_\infty > \hat{c}_\alpha - n^{-1}\big) - \alpha - \sup_{z\ge 0}\big|\mathbb{P}_g\big(|\mathbf{G}^*|_\infty \le z\big) - \mathbb{P}\big(|\mathbf{G}|_\infty \le z\big)\big| \\
&\qquad - \sup_{z\ge 0}\big|\mathbb{P}\big(|\mathbf{G}|_\infty \le z\big) - \mathbb{P}_{H_0}\big(\widehat{T}_{\max} \le z\big)\big| - Cn^{-1}(\log p)^{1/2} \\
&> - \sup_{z\ge 0}\big|\mathbb{P}_g\big(|\mathbf{G}^*|_\infty \le z\big) - \mathbb{P}\big(|\mathbf{G}|_\infty \le z\big)\big| \\
&\qquad - \sup_{z\ge 0}\big|\mathbb{P}\big(|\mathbf{G}|_\infty \le z\big) - \mathbb{P}_{H_0}\big(\widehat{T}_{\max} \le z\big)\big| - Cn^{-1}(\log p)^{1/2}.
\end{aligned}
$$

Combining the above calculations completes the proof of Theorem 1. □

B.2 Proof of Theorem 2

Since $\bar{p} = p(p+1)/2 \le p^2$, we have

$$
\{2\log(\bar{p})\}^{1/2} + \{2\log(1/\alpha)\}^{1/2} \le 2(\log p)^{1/2} + \{2\log(1/\alpha)\}^{1/2},
$$

Employing arguments similar to those used in the proof of Theorem 4 in Chang et al. (2014) by taking $p = \bar{p}$, $\mu_{1,a} = \sigma_{1,u_a v_a}$ and $\mu_{2,a} = \sigma_{2,u_a v_a}$ for $a = 1, \ldots, \bar{p}$, the conclusion follows immediately from Lemma A.6. □

C More numerical results

C.1 Additional simulation results on the proposed testing procedure

In this section, we include some extra simulation results for comparing the empirical sizes and powers of the proposed test $\Psi_{B,\alpha}$ with existing methods in literature. We generated two independent random samples $\{\mathbf{X}_i\}_{i=1}^n$ and $\{\mathbf{Y}_j\}_{j=1}^m$ from multivariate models $\mathbf{X}_i = \mathbf{\Sigma}_{1,*}^{1/2}\mathbf{Z}_i^{(1)}$ and $\mathbf{Y}_j = \mathbf{\Sigma}_{2,*}^{1/2}\mathbf{Z}_j^{(2)}$, where $Z_{ik}^{(1)}$ and $Z_{jk}^{(2)}$ are negative binomial random variables such that $Z_{ik}^{(1)} \sim \mathrm{NB}(100, 0.2)$ and $5^{-1/2}Z_{jk}^{(2)} \sim \mathrm{NB}(200, 0.5)$. The four covariance structures M1–M4

Table S1: Empirical sizes of the proposed test $\Psi_{B,\alpha}$ along with those of the tests by Li and Chen (2012) (LC), Schott (2007) (Sc), and Cai et al. (2013) (CLX) for data distributed according to the negative binomial distribution as discussed above. The covariance structures M1–M4 are discussed in Section 3 in the paper.

(n_1, n_2)	p	80	280	500	1000	80	280	500	1000
		Covariance structure M1				Covariance structure M2			
$(45, 45)$	$\Psi_{B,\alpha}$	0.078	0.078	0.093	0.090	0.076	0.071	0.090	0.094
	LC	0.073	0.061	0.055	0.052	0.080	0.073	0.080	0.082
	Sc	0.066	0.054	0.046	0.047	0.063	0.067	0.067	0.067
	CLX	0.071	0.053	0.065	0.062	0.064	0.054	0.069	0.060
$(60, 80)$	$\Psi_{B,\alpha}$	0.066	0.064	0.067	0.075	0.069	0.070	0.071	0.072
	LC	0.060	0.064	0.058	0.056	0.059	0.099	0.071	0.087
	Sc	0.055	0.057	0.054	0.050	0.057	0.092	0.072	0.068
	CLX	0.051	0.065	0.055	0.059	0.054	0.062	0.054	0.057
		Covariance structure M3				Covariance structure M4			
$(45, 45)$	$\Psi_{B,\alpha}$	0.077	0.065	0.084	0.091	0.065	0.078	0.084	0.084
	LC	0.081	0.065	0.057	0.051	0.061	0.051	0.047	0.035
	Sc	0.061	0.051	0.047	0.035	0.059	0.049	0.046	0.052
	CLX	0.065	0.040	0.071	0.084	0.050	0.059	0.061	0.089
$(60, 80)$	$\Psi_{B,\alpha}$	0.064	0.067	0.092	0.065	0.052	0.082	0.061	0.081
	LC	0.057	0.053	0.060	0.057	0.050	0.060	0.065	0.046
	Sc	0.054	0.051	0.054	0.048	0.044	0.051	0.058	0.037
	CLX	0.053	0.059	0.075	0.043	0.044	0.070	0.051	0.064

in Section 3 were considered and $\mathbf{\Sigma}_{1,*} = \mathbf{\Sigma}_{2,*} = \mathbf{\Sigma}_*$ under the null hypothesis in (2.1) in the paper. The empirical sizes and powers are displayed in Table S1 and Figure S1, respectively. For power comparisons, the alternatives were set in the same way as those in Section 3. Results are based on 1000 replications with $\alpha = 0.05$.

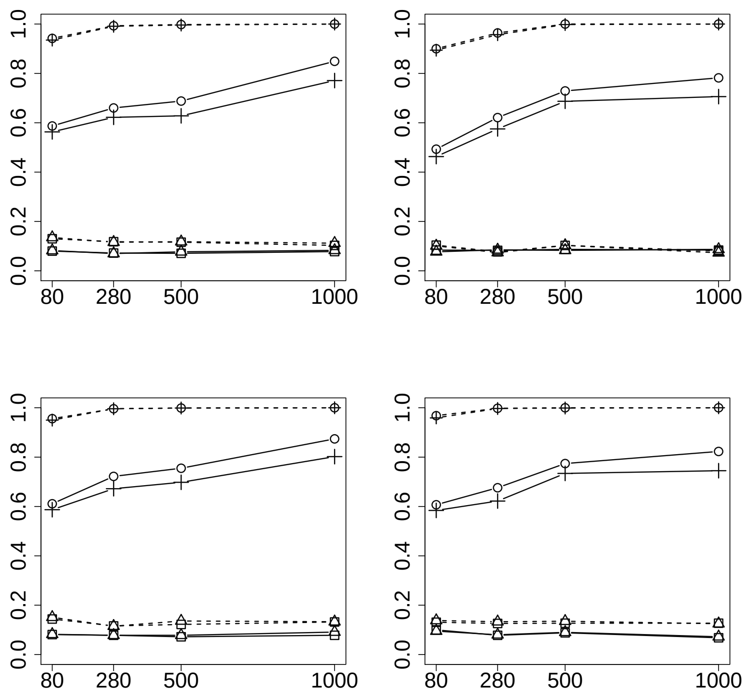

Figure S1: Comparison of empirical powers for data generated by data models with the negative binomial distribution and covariance structures M1–M4. In each panel, horizontal and vertical axes depict dimension p and empirical powers, respectively; and unbroken lines and dashed lines represent the results for $(n_1, n_2) = (45, 45)$ and $(60, 80)$, respectively. The different symbols on the lines represent different tests experimented in the study, where $\circ$, $\square$, $\triangle$, and $+$ indicate the proposed test $\Psi_{B,\alpha}$, tests by Li and Chen (2012), Schott (2007), and Cai et al. (2013), respectively. Results are based on 1000 replications with $\alpha = 0.05$.

C.2 Numerical results on the proposed clustering procedure

First, we demonstrate the performance of the proposed dissimilarity measure and the clustering procedure in Figure S2. As discussed by Hastie, Tibshirani and Friedman (2010), clustering based on correlation (similarity) is equivalent to that based on the squared Euclidean distance (dissimilarity). By comparing panels (b) and (c), we observe that the sample correlation matrix is masked by noises while the proposed measure $d_{k\ell}$ reasonably captures the similarities among variables. Therefore, in this example, the proposed dissimilarity performs better than the traditional Euclidean distance as a dissimilarity measure for variable clustering.

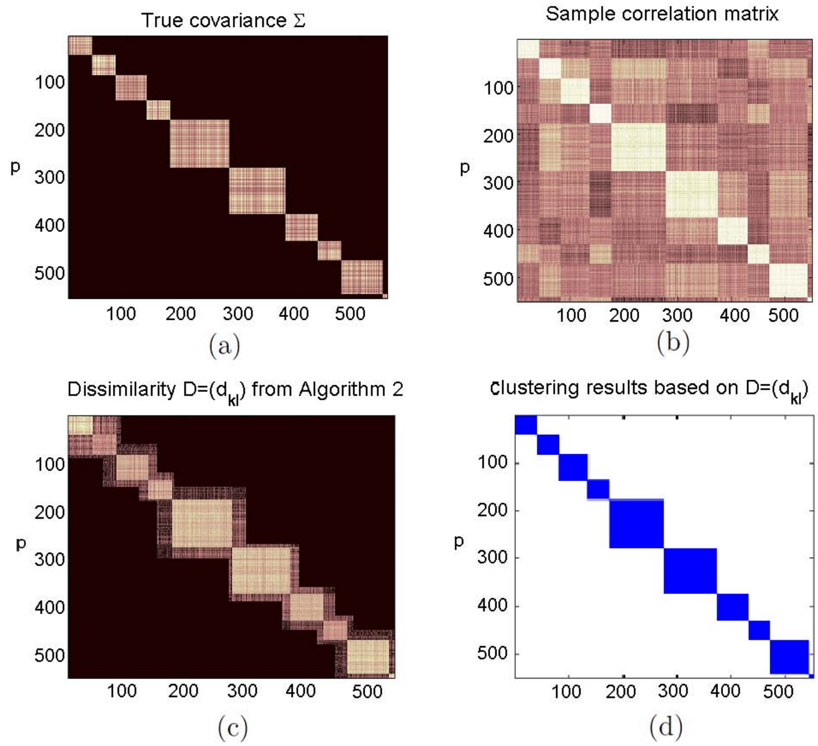

Figure S2: The true $p \times p$ covariance matrix $\mathbf{\Sigma}_1$ with $p = 550$, as displayed in panel (a), is in block-diagonal with 10 blocks in different sizes. Entries within each blocks are generated by $\sigma_{1,k\ell} \overset{\text{i.i.d.}}{\sim} \text{Unif}(1.2\sqrt{(\log p)/n}, 1)$ for $k \neq \ell$ and $\sigma_{1,kk} \overset{\text{i.i.d.}}{\sim} \text{Unif}(0.5, 2.5)$, where $n = 30$. Entries across blocks are equal to zero. Using data generation model (D1) in Section 3 in conjunction with such a $\mathbf{\Sigma}_1$, 30 independent samples are obtained whose sample correlation matrix is displayed in panel (b). The dissimilarity matrix $\mathbf{D} = (d_{k\ell})$ in (4.2) is computed based on these 30 independent samples and displayed in panel (c) that the light pixel indicates $d_{k\ell}$ is close to one. The clustering results of p variables based on the proposed dissimilarity $\mathbf{D}$ is shown in panel (d). This figure appears in color in the electronic version of this article.

We further investigate the proposed dissimilarity measure and compare it to alternatives using the asthma data (Voraphani et al., 2014). We considered different clustering methods, including K-mean and the hierarchical clustering algorithm with average linkage, in conjunction with different dissimilarity measures, including the Euclidean distance, the cityblock measure (which is the sum of absolute differences), the Chebychev dissimilarity (which is the maximum coordinate difference), and the correlation dissimilarity (which is one minus the sample correlation magnitude). We compared the numbers of clusters among genes from both the health and asthma groups determined by the proposed measure $d_{k\ell}$ to those determined by the aforementioned traditional dissimilarity measures. For the K-mean method, we employed the Gap statistic to determine the number of clusters; while for the hierarchical algorithm, we used the Calinski-Harabasz criterion. We considered the GO terms analyzed in Section 5 in the paper, which are GO terms 0032689 (negative regulation of IFN-γ production), 0060333 (IFN-γ-mediated signaling pathway) and 0071346 (cellular response to IFN-γ), 0004601 (peroxidase activity), 0042446 (hormone biosynthetic process), 0035162 (embryonic hemopoiesis), 0006979 (response to oxidative stress), and 0009986 (cell surface). The results are summarized in Tables S2 and S3 below.

As studied in the paper, these GO terms all have different covariance structures between the health and asthma groups. Naturally, their gene clustering results are expected to be different between the two groups. From the results based on the hierarchical algorithm displayed in Table S2, in general, the numbers of clusters determined by the traditional dissimilarity measures are very similar between the health and asthma groups for each GO term and are fairly large compared to those determined by the proposed measure $d_{k\ell}$. For some GO terms, the numbers of clusters determined by the traditional dissimilarity measures are even close to the size of the GO terms (for example, GO:0042446, GO:0035162, and GO:0071346), which suggests a lack of differentiation in gene clustering structures between the health and asthma groups. Similar observations are made from the results based on the K-means method in Table S3. This contradicts with the statistical conclusions on differential covariance structures as well as the biological knowledge regarding those GO terms (Voraphani et al., 2014). On the other hand, we also display the clustering results for GO:0004601 from the K-mean method based on the Euclidean and cityblock distances in Figures S3 and S4. As discussed in the paper, the proposed dissimilarity measure is able to discover the biologically interesting differential gene clusters including, for example, those regarding DUOX2 in GO:0004601. From Figures S3 and S4, the two traditional measures are not successful in capturing the large differential clustering structures within GO:0004601 between the health and asthma groups. Such a differential clustering structure is, however, important for asthma as discussed by Voraphani et al. (2014).

Table S2: Comparison of the numbers of clusters determined by the proposed dissimilarity measure $d_{k\ell}$ with those determined by different dissimilarity measures, including the Euclidean distance, cityblock dissimilarity and Chebychev dissimilarity, based on the hierarchical clustering algorithm with average linkage. In the parenthesis, it reports the numbers of clusters among genes within each GO term for the health and asthma groups, respectively.

	GO term size	Euclidean	cityblock	Chebychev	ours
GO:0071346	13	(11,11)	(11,11)	(11,11)	(9,5)
GO:0032689	17	(5,5)	(5,5)	(6,5)	(16,13)
GO:0035162	29	(23,23)	(21,23)	(21,23)	(4,4)
GO:0004601	30	(23,23)	(23,23)	(19,23)	(19,9)
GO:0042446	39	(33,33)	(5,33)	(37,33)	(7,25)
GO:0060333	130	(107,107)	(83,105)	(3,3)	(4,92)
GO:0009986	439	(9,7)	(7,7)	(5,5)	(4,14)

Table S3: Comparison of the numbers of clusters determined by the proposed dissimilarity measure $d_{k\ell}$ with those determined by different dissimilarity measures, including the Euclidean distance, cityblock dissimilarity and correlation, based on the K-mean algorithm. In the parenthesis, it reports the numbers of clusters among genes within each GO term for the health and asthma groups, respectively.

	GO term size	Euclidean	cityblock	correlation	ours
GO:0071346	13	(1,1)	(1,9)	(1,1)	(9,5)
GO:0035162	29	(19,19)	(19,13)	(1,1)	(4,4)
GO:0004601	30	(18,18)	(14,15)	(1,1)	(19,9)
GO:0042446	39	(17,25)	(21,23)	(1,1)	(7,25)
GO:0060333	130	(87,83)	(81,87)	(85,80)	(4,92)
GO:0006979	157	(103,93)	(93,101)	(1,105)	(107,7)
GO:0009986	439	(263,279)	(287,285)	(273,291)	(4,14)

D Real data analysis

In this section, we applied the procedure proposed by Chang et al. (2014) to test the global hypotheses

$$H_{0g}^m : \boldsymbol{\mu}_{h,g} = \boldsymbol{\mu}_{a,g} \quad \text{versus} \quad H_{1g}^m : \boldsymbol{\mu}_{h,g} \neq \boldsymbol{\mu}_{a,g} \tag{B.1}$$

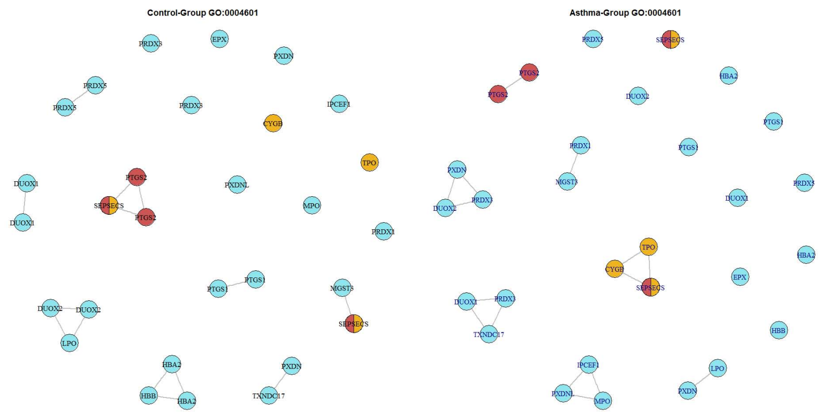

Figure S3: Comparison of clustering structures of GO:0004601, peroxidase activity, between health and disease groups using K-mean clustering with Euclidean distance. The number of clusters is determined using the Gap statistic. This figure appears in color in the electronic version of this article.

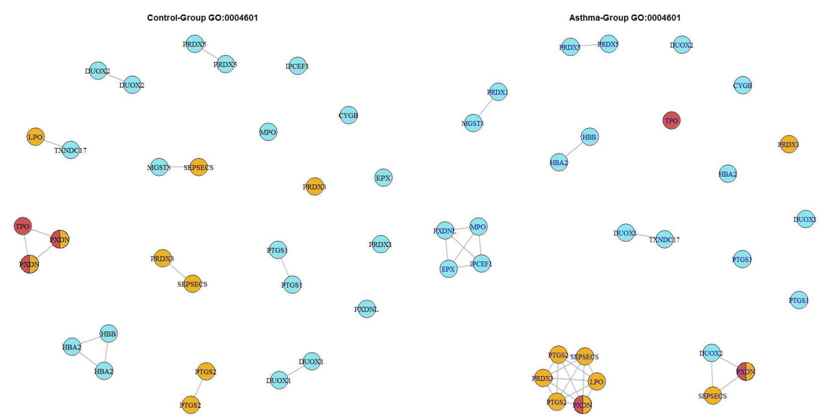

Figure S4: Comparison of clustering structures of GO:0004601, peroxidase activity, between health and disease groups using K-mean clustering with cityblock distance, i.e. the sum of absolute differences. The number of clusters is determined using the Gap statistic. This figure appears in color in the electronic version of this article.

as a comparison. By controlling the FDR at 2.5%, testing the mean hypotheses (B.1) declared $2,122$ GO terms significant. Besides the 701 GO terms that were declared significant via testing both (5.1) and (B.1), the proposed test on intergene relationships discovered 268 significant GO terms that were not identified by the traditional differential analysis based on the mean test. This also reflects the lately growing demands on analysing gene dependence structures separately for analysing mean gene expressions. Table S4 displays the top 15 most significant GO terms declared by $\Psi_{B,\alpha}$ while undetected via testing (B.1).

Table S4: Top 15 most significant GO terms declared by the proposed test $\Psi_{B,\alpha}$ for testing (5.1) with FDR controlled at level 2.5% but not identified by testing (B.1).

GO ID	GO term name
GO:0060992	response to fungicide
GO:0017187	peptidyl-glutamic acid carboxylation
GO:0034451	centriolar satellite
GO:0007548	sex differentiation
GO:0019902	phosphatase binding
GO:0050771	negative regulation of axonogenesis
GO:0006298	mismatch repair
GO:0030983	mismatched DNA binding
GO:0046856	phosphatidylinositol dephosphorylation
GO:0048854	brain morphogenesis
GO:0050427	3'-phosphoadenosine 5'-phosphosulfate metabolic process
GO:0045907	positive regulation of vasoconstriction
GO:0010667	negative regulation of cardiac muscle cell apoptosis
GO:0048037	cofactor binding
GO:0018107	peptidyl-threonine phosphorylation

Also, we include more results in Figures S5–S9 from the case study on human severe asthma dataset in Section 5.

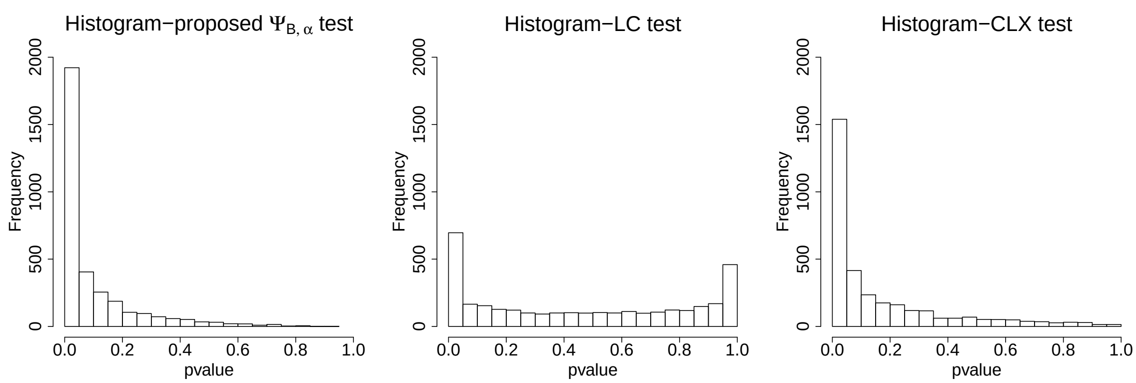

Figure S5: Histograms of p-values of the proposed test $\Psi_{B,\alpha}$ along with those of the CLX and the LC tests using the human severe asthma dataset with $3,290$ GO terms.

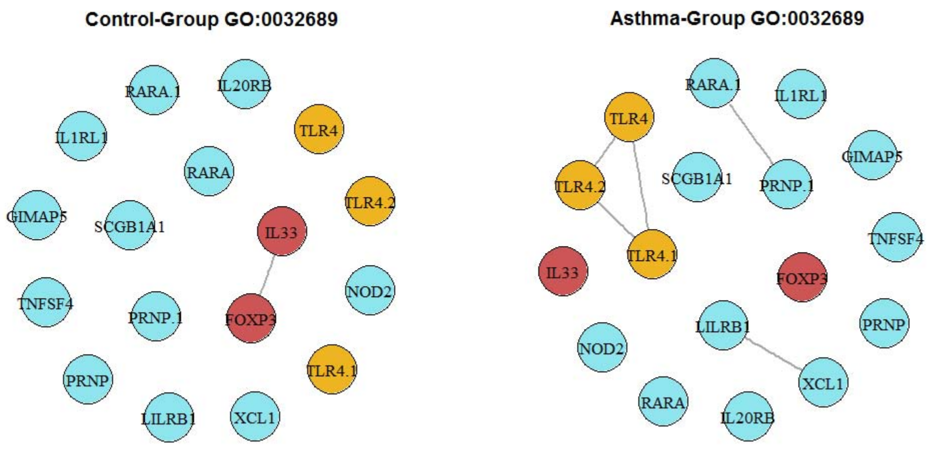

Figure S6: Comparison of clustering structures of GO:0032689, negative regulation of IFN-γ production, between the health and disease groups using the proposed clustering procedure. This figure appears in color in the electronic version of this article.

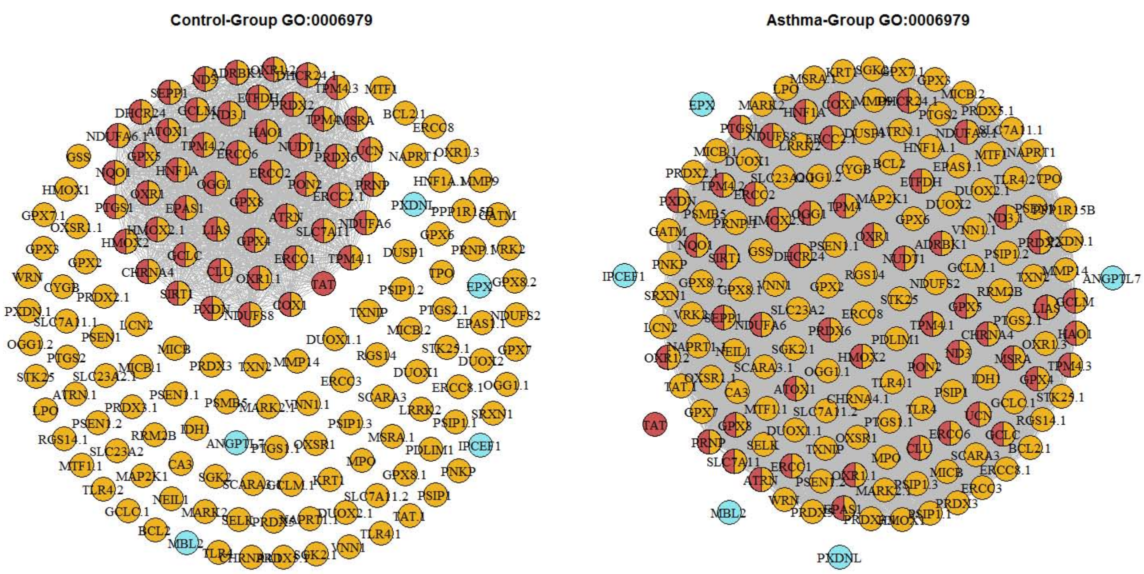

Figure S7: Comparison of clustering structures of GO:0006979, response to oxidative stress, between the health and disease groups using the proposed clustering procedure. This figure appears in color in the electronic version of this article.

References

Cai, T. T., Liu, W., and Xia, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association* **108**, 265–277.

Chang, J., Zhou, W., and Zhou, W.-X. (2014). Simulation-based hypothesis testing of high

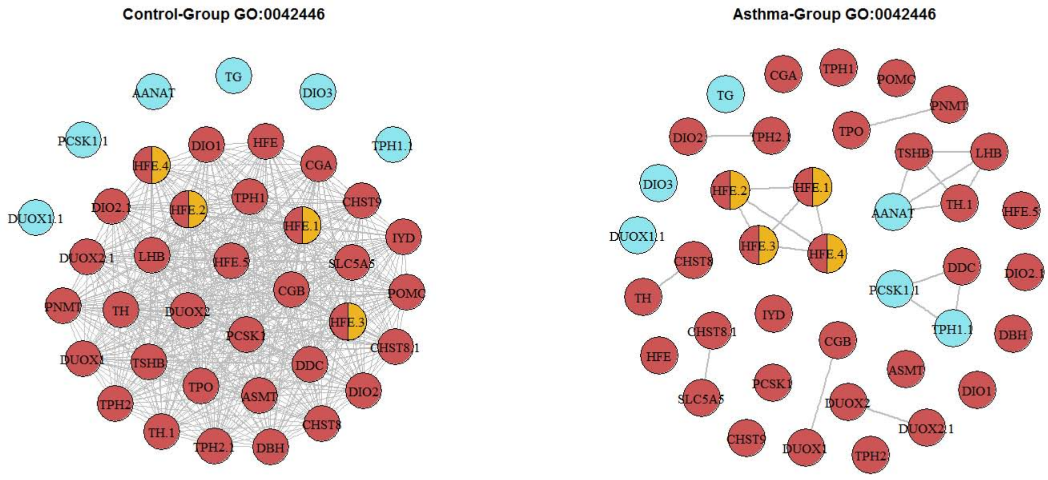

Figure S8: Comparison of clustering structures of GO:0042446, hormone biosynthetic process, between the health and disease groups using the proposed clustering procedure. This figure appears in color in the electronic version of this article.

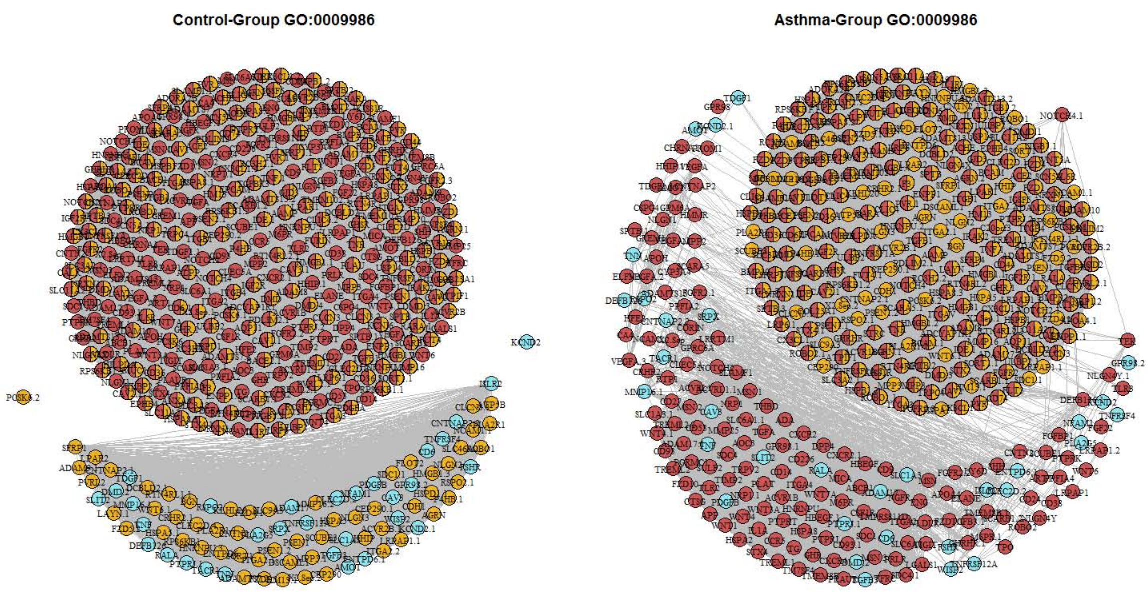

Figure S9: Comparison of clustering structures of GO:0009986, cell surface, between the health and disease groups using the proposed clustering procedure. This figure appears in color in the electronic version of this article.

dimensional means under covariance heterogeneity. Available at *arXiv:1406.1939.*

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics* **41**, 2786–2819.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2015). Comparison and anticoncentration bounds for maxima of Gaussian random vectors. *Probability Theory and Related Fields* **162**, 47–70.

de la Peña, V. H., Lai, T. L., and Shao, Q.-M. (2009). *Self-Normalized Processes: Limit Theory and Statistical Applications.* Berlin: Springer.

Delaigle, A., Hall, P., and Jin, J. (2011). Robustness and accuracy of methods for high dimensional data analysis based on Student's t-Statistic. *Journal of the Royal Statistical Society, Series B* **73**, 283–301.

Einmahl, U. and Li, D. (2008). Characterization of LIL behavior in Banach space. *Transactions of the American Mathematical Society* **360**, 6677–6693.

Hastie, T., Tibshirani, R., and Friedman, J. (2010). *The Elements of Statistical Learning.* 2nd edition. Springer-Verlag: New York.

Li, J. and Chen, S. X. (2012). Two-sample tests for high-dimensional covariance matrices. *The Annals of Statistics* **40**, 908–940.

Schott, J. R. (2007). A test for the equality of covariance matrices when the dimension is large relative to the sample size. *Computational Statistics and Data Analysis* **51**, 6535–6542.

Voraphani, N., Gladwin, M. T., Contreras, A. U., Kaminski, N., Tedrow, J. R., Milosevic, J., Bleecker, E. R., Meyers, D. A., Ray, A., Ray, P., Erzurum, S. C., Busse, W. W., Zhao, J., Trudeau, J. B., and Wenzel, S. E. (2014). An airway epithelial iNOS-DUOX2-thyroid peroxidase metabolome drives Th1/Th2 nitrative stress in human severe asthma. *Mucosal Immunology* **7**, 1175–1185.